Optimal and Better Transport Plans

Mathias Beiglböck [a,1], Martin Goldstern [a], Gabriel Maresch [b,2], Walter Schachermayer [b,3]

[a] *Institut für Diskrete Mathematik und Geometrie, Technische Universität Wien Wiedner Hauptstraße 8-10/104,1040 Wien, Austria*

[b] *Institut für Wirtschaftsmathematik, Technische Universität Wien Wiedner Hauptstraße 8-10/105, 1040 Wien, Austria*

Abstract

We consider the Monge-Kantorovich transport problem in a purely measure theoretic setting, i.e. without imposing continuity assumptions on the cost function. It is known that transport plans which are concentrated on c-monotone sets are optimal, provided the cost function c is either lower semi-continuous and finite, or continuous and may possibly attain the value ∞. We show that this is true in a more general setting, in particular for merely Borel measurable cost functions which are finite almost everywhere on an open set. In a previous paper Schachermayer and Teichmann considered strongly c-monotone transport plans and proved that every strongly c-monotone transport plan is optimal. We establish necessary and sufficient conditions on c-monotone transport plans to be strongly c-monotone.

Key words: Monge-Kantorovich problem, c-cyclically monotone, strongly c-monotone, measurable cost function

1 Introduction

We consider the *Monge-Kantorovich transport problem* (μ, ν, c) for Borel probability measures μ, ν on Polish spaces X, Y and a Borel measurable cost function $c : X \times Y \to [0, \infty]$. As standard references on the theory of mass transport

[1] Supported by the Austrian Science Fund (FWF) under grant S9612.
[2] Supported by the Austrian Science Fund (FWF) under grant SY328.
[3] Supported by the Austrian Science Fund (FWF) under grant P19456, from the Vienna Science and Technology Fund (WWTF) under grant MA13 and from the Christian Doppler Research Association (CDG)

we mention [1,8,13,14]. By $\Pi(\mu,\nu)$ we denote the set of all probability measures on $X \times Y$ with X-marginal μ and Y-marginal ν. For a Borel measurable *cost function* $c : X \times Y \to [0,\infty]$ the transport costs of a given transport plan π are defined by

$$I_c[\pi] := \int_{X\times Y} c(x,y)d\pi. \tag{1}$$

π is called a *finite* transport plan if $I_c[\pi] < \infty$.

A nice interpretation of the Monge-Kantorovich transport problem is given by Cédric Villani in Chapter 3 of the impressive monograph [14]:

> "Consider a large number of bakeries, producing breads, that should be transported each morning to cafés where consumers will eat them. The amount of bread that can be produced at each bakery, and the amount that will be consumed at each café are known in advance, and can be modeled as probability measures (there is a "density of production" and a "density of consumption") on a certain space, which in our case would be Paris (equipped with the natural metric such that the distance between two points is the length of the shortest path joining them). The problem is to find in practice where each unit of bread should go, in such a way as to minimize the total transport cost."

We are interested in *optimal* transport plans, i.e. minimizers of the functional $I_c[\cdot]$ and their characterization via the notion of c-monotonicity.

Definition 1.1 *A Borel set* $\Gamma \subseteq X \times Y$ *is called* c-monotone *if*

$$\sum_{i=1}^{n} c(x_i, y_i) \leq \sum_{i=1}^{n} c(x_i, y_{i+1}) \tag{2}$$

for all pairs $(x_1, y_1), \ldots, (x_n, y_n) \in \Gamma$ *using the convention* $y_{n+1} := y_1$*. A transport plan* π *is called c-monotone if there exists a c-monotone* Γ *with* $\pi(\Gamma) = 1$*.*

In the literature (e.g. [1,2,6,7,12]) the following characterization was established under various continuity assumptions on the cost function. Our main result states that those assumptions are not required.

Theorem 1 *Let* X, Y *be Polish spaces equipped with Borel probability measures* μ, ν *and let* $c : X \times Y \to [0,\infty]$ *a Borel measurable cost function.*

a. Every finite optimal transport plan is c-monotone.
b. Every finite c-monotone transport plan is optimal if there exist a closed set F *and a* $\mu \otimes \nu$*-null set* N *such that* $\{(x,y) : c(x,y) = \infty\} = F \cup N$*.*

Thus in the case of a cost function which does not attain the value ∞ the equivalence of optimality and c-monotonicity is valid without any restrictions beyond the obvious measurability conditions inherent in the formulation of the problem.

The subsequent construction due to Ambrosio and Pratelli in [1, Example 3.5] shows that if c is allowed to attain ∞ the implication "c-monotone $\Rightarrow$ optimal" does not hold without some additional assumption as in Theorem 1.b.

Example 1.2 (Ambrosio and Pratelli) *Let $X = Y = [0,1]$, equipped with Lebesgue measure $\lambda = \mu = \nu$. Pick $\alpha \in [0,1)$ irrational. Set*

$$\Gamma_0 = \{(x,x) : x \in X\}, \quad \Gamma_1 = \{(x, x \oplus \alpha) : x \in X\},$$

where $\oplus$ is addition modulo 1. Let $c : X \times Y \to [0,\infty]$ be such that $c = a \in [0,\infty)$ on Γ_0, $c = b \in [0,\infty)$ on Γ_1 and $c = \infty$ otherwise. It is then easy to check that Γ_0 and Γ_1 are c-monotone sets. Using the maps $f_0, f_1 : X \to X \times Y$, $f_0(x) = (x,x), f_1(x) = (x, x \oplus \alpha)$ one defines the transport plans $\pi_0 = f_{0\#}\lambda$, $\pi_1 = f_{1\#}\lambda$ supported by Γ_0 respectively Γ_1. Then π_0 and π_1 are finite c-monotone transport plans, but as $I_c[\pi_0] = a, I_c[\pi_1] = b$ it depends on the choice of a and b which transport plan is optimal. Note that in contrast to the assumption in Theorem 1.b the set $\{(x,y) \in X \times Y : c = \infty\}$ is open.

We want to remark that rather trivial (folkloristic) examples show that no optimal transport has to exist if the cost function doesn't satisfy proper continuity assumptions.

Example 1.3 *Consider the task to transport points on the real line (equipped with the Lebesgue measure) from the interval $[0,1)$ to $[1,2)$ where the cost of moving one point to another is the squared distance between these points ($X = [0,1), Y = [1,2)$, $c(x,y) = (x-y)^2$, $\mu = \nu = \lambda$). The simplest way to achieve this transport is to shift every point by 1. This results in transport costs of 1 and one easily checks that all other transport plans are more expensive.*

If we now alter the cost function to be 2 whenever two points have distance 1, i.e. if we set

$$\tilde{c}(x,y) = \begin{cases} 2 & \text{if } y = x+1 \\ c(x,y) & \text{otherwise} \end{cases},$$

it becomes impossible to find a transport plan $\pi \in \Pi(\mu,\nu)$ with total transport costs $I_{\tilde{c}}[\pi] = 1$, but it is still possible to achieve transport costs arbitrarily close to 1. (For instance, shift $[0, 1-\varepsilon)$ to $[1+\varepsilon, 2)$ and $[1-\varepsilon, 1)$ to $[1, 1+\varepsilon)$ for small $\varepsilon > 0$.) Note that the dual optimizers $\varphi(x) = -1 - 2x, \psi(y) = 2y$ *are not affected by this change of the cost function.*

The notion of c-monotonicity originates in convex analysis. The well known Rockafellar Theorem (see for instance [10, Theorem 3] or [13, Theorem 2.27]) and its generalization, Rüschendorf's Theorem (see [11, Lemma 2.1]), characterize c-monotonicity in $\mathbb{R}^n$ in terms of integrability. The definitions of c-concave functions and super-differentials can be found for instance in [13, Section 2.4].

Theorem (Rockafellar) *A nonempty set $\Gamma \subseteq \mathbb{R}^n \times \mathbb{R}^n$ is cyclically monotone (that is, c-monotone with respect to the squared euclidean distance) if and only if there exists a l.s.c. concave function $\varphi : \mathbb{R}^n \to \mathbb{R}$ such that Γ is contained in the super-differential $\partial(\varphi)$.*

Theorem (Rüschendorf) *Let X, Y be abstract spaces and $c : X \times Y \to [0, \infty]$ arbitrary. Let $\Gamma \subseteq X \times Y$ be c-monotone. Then there exists a c-concave function $\varphi : X \to Y$ such that Γ is contained in the c-super-differential $\partial^c(\varphi)$.*

Important results of Gangbo and McCann [2] and Brenier [13, Theorem 2.12] use these potentials to establish uniqueness of the solutions of the Monge-Kantorovich transport problem in $\mathbb{R}^n$ for different types of cost functions subject to certain regularity conditions.

Optimality implies c-monotonicity: This is evident in the discrete case if X and Y are finite sets. For suppose that π is a transport plan for which c-monotonicity is violated on pairs $(x_1, y_1), \dots, (x_n, y_n)$ where all points $x_1, \dots, x_n$ and $y_1, \dots, y_n$ carry positive mass. Then we can reduce costs by sending the mass $\alpha > 0$, for α sufficiently small, from x_i to y_{i+1} instead of y_i, that is, we replace the original transport plan π with

$$\pi^\beta = \pi + \alpha \sum_{i=1}^{n} \delta_{(x_i, y_{i+1})} - \alpha \sum_{i=1}^{n} \delta_{(x_i, y_i)}. \tag{3}$$

(Here we are using the convention $y_{n+1} = y_1$.)

Gangbo and McCann ([2, Theorem 2.3]) show how continuity assumptions on the cost function can be exploited to extend this to an abstract setting. Hence one achieves:

Let X and Y be Polish spaces equipped with Borel probability measures μ, ν. Let $c : X \times Y \to [0, \infty]$ be a l.s.c. cost function. Then every finite optimal transport plan is c-monotone.

Using measure theoretic tools, as developed in the beautiful paper by Kellerer [5], we are able to extend this to Borel measurable cost functions (Theorem

1.a.) without any additional regularity assumption.

c-monotonicity implies optimality: In the case of finite spaces X, Y this again is nothing more than an easy exercise ([13, Exercise 2.21]). The problem gets harder in the infinite setting. It was first proved in [2] that for X, Y compact subsets of $\mathbb{R}^n$ and c a continuous cost function, c-monotonicity implies optimality. In a more general setting this was shown in [1, Theorem 3.2] for l.s.c. cost functions which additionally satisfy the moment conditions

$$\mu\left(\left\{x : \int_Y c(x,y)d\nu < \infty\right\}\right) > 0,$$
$$\nu\left(\left\{y : \int_X c(x,y)d\mu < \infty\right\}\right) > 0.$$

Further research into this direction was initiated by the following problem posed by Villani in [13, Problem 2.25]:

For $X = Y = \mathbb{R}^n$ and $c(x,y) = \|x-y\|^2$, the squared euclidean distance, does c-monotonicity of a transport plan imply its optimality?

A positive answer to this question was given independently by Pratelli in [7] and by Schachermayer and Teichmann in [12]. Pratelli proves the result for countable spaces and shows that it extends it to the Polish case by means of approximation if the cost function $c : X \times Y \to [0, \infty]$ is continuous. The paper [12] pursues a different approach: The notion of *strong c-monotonicity* is introduced. From this property optimality follows fairly easily and the main part of the paper is concerned with the fact that strong c-monotonicity follows from the usual notion of c-monotonicity in the Polish setting if c is assumed to be l.s.c. and finitely valued.

Part (b) of Theorem 1 unifies these statements: Pratelli's result follows from the fact that for continuous $c : X \times Y \to [0, \infty]$ the set $\{c = \infty\} = c^{-1}[\{\infty\}]$ is closed; the Schachermayer-Teichmann result follows since for finite c the set $\{c = \infty\}$ is empty.

Similar to [12] our proofs are based on the concept of *strong c-monotonicity.* In Section 1.2 we present *robust optimality* which is a variant of optimality that we shall show to be equivalent to strong c-monotonicity. As not every optimal transport plan is also robustly optimal, this accounts for the somewhat provocative concept of "better than optimal" transport plans alluded to in the title of this paper.

Correspondingly the notion of strong c-monotonicity is in fact stronger than ordinary c-monotonicity (at least if c is allowed to assume the value ∞).

It will be of fundamental importance to give a characterization of strong c-monotonicity in terms of ordinary c-monotonicity. This connection is explained in detail in Section 1.3 below.

1.2 Strong Notions

It turns out that optimality of a transport plan is intimately connected with the notion of *strong c-monotonicity* introduced in [12].

Definition 1.4 *A Borel set* $\Gamma \subseteq X \times Y$ *is* strongly c-monotone *if there exist Borel measurable functions* $\varphi : X \to \mathbb{R} \cup \{-\infty\}$ *and* $\psi : Y \to \mathbb{R} \cup \{-\infty\}$ *such that* $\varphi(x) + \psi(y) \leq c(x, y)$ *for all* $(x, y) \in X \times Y$ *and* $\varphi(x) + \psi(y) = c(x, y)$ *for all* $(x, y) \in \Gamma$*. A transport plan* $\pi \in \Pi(\mu, \nu)$ *is* strongly c-monotone *if* π *is concentrated on a strongly c-monotone Borel set* Γ*.*

Strong c-monotonicity implies c-monotonicity since

$$\sum_{i=1}^{n} c(x_{i+1}, y_i) \geq \sum_{i=1}^{n} \varphi(x_{i+1}) + \psi(y_i) = \sum_{i=1}^{n} \varphi(x_i) + \psi(y_i) = \sum_{i=1}^{n} c(x_i, y_i) \tag{4}$$

whenever $(x_1, y_1), \ldots, (x_n, y_n) \in \Gamma$.

If there are *integrable* functions φ and ψ witnessing that π is strongly c-monotone, then for every $\tilde{\pi} \in \Pi(\mu, \nu)$ we can estimate:

$$\begin{aligned} I_c[\pi] = \int_\Gamma c(x, y) d\pi &= \int_\Gamma [\varphi(x) + \psi(y)] d\pi = \\ = \int_\Gamma \varphi(x) d\mu + \int_\Gamma \psi(y) d\nu &= \int_\Gamma [\varphi(x) + \psi(y)] d\tilde{\pi} \leq I_c[\tilde{\pi}]. \end{aligned}$$

Thus in this case strong c-monotonicity implies optimality. However there is no reason why the Borel measurable functions φ, ψ appearing in Definition 1.4 should be integrable. In [12, Proposition 2.1] it is shown that, for l.s.c. cost functions, there is a way of truncating which allows to also handle non-integrable functions φ and ψ. An inspection of the proof reveals that it uses only Borel measurability of the cost function, so we may rephrase [12, Proposition 2.1] in the following form:

Let X, Y *be Polish spaces equipped with Borel probability measures* μ, ν *and let* $c : X \times Y \to [0, \infty]$ *be Borel measurable. Then every finite transport plan which is strongly c-monotone is optimal.*

As it will turn out, strongly c-monotone transport plans even satisfy a "better" notion of optimality, called *robust optimality.*

Definition 1.5 *Let X, Y be Polish spaces equipped with Borel probability measures μ, ν and let $c : X \times Y \to [0, \infty]$ be a Borel measurable cost function. A transport plan $\pi \in \Pi(\mu, \nu)$ is* robustly optimal *if, for any Polish space Z and any finite Borel measure λ on Z, there exists a Borel measurable extension $\tilde{c} : (X \cup Z) \times (Y \cup Z) \to [0, \infty]$ satisfying*

$$\tilde{c}(a, b) = \begin{cases} c(a, b) & \text{for} \quad a \in X, b \in Y \\ 0 & \text{for} \quad a, b \in Z \\ < \infty & \text{otherwise} \end{cases}$$

such that the measure $\tilde{\pi} := \pi + (id_Z \times id_Z)_{\#}\lambda$ is optimal on $(X \cup Z) \times (Y \cup Z)$. Note that $\tilde{\pi}$ is not a probability measure, but has total mass $1 + \lambda(Z) \in [1, \infty)$

Robust optimality has a colorful "economic" interpretation: a tycoon wants to enter the Parisian croissant consortium. She builds a storage of size $\lambda(Z)$ where she buys up croissants and sends them to the cafés. Her hope is that by offering low transport costs, the previously optimal transport plan π will not be optimal anymore, so that the traditional relations between bakeries and cafés will collapse. Of course, the authorities of Paris will try to defend their structure by imposing (possibly very high, but still finite) tolls for all transports to and from the tycoon's storage, thus resulting in finite $\tilde{c}(a, b)$ for $(a, b) \in (X \times Z) \cup (Z \times Y)$. In the case of robustly optimal π they can successfully defend themselves against the intruder.

Clearly every robustly optimal transport π plan is optimal in the usual sense and therefore c-monotone. The crucial feature is that robust optimality implies strong c-monotonicity. In fact, the two properties are equivalent.

Theorem 2 *Let X, Y be Polish spaces equipped with Borel probability measures μ, ν and $c : X \times Y \to [0, \infty]$ a Borel measurable cost function. For a finite transport plan π the following assertions are equivalent:*

a. π is strongly c-monotone.
b. π is robustly optimal.

Example 5.3 below shows that robust optimality resp. strong c-monotonicity is in fact a stronger property than usual optimality.

1.3 Relating c-monotonicity and strong c-monotonicity

A c-monotone transport plan resists the attempt of enhancement by means of cyclically rerouting in the spirit of Proposition 1.1. This, however, may be due to the fact that rerouting is a priori impossible due to infinite transport costs on certain routes.

Continuing Villani's interpretation, a situation where rerouting in this consortium of bakeries and cafés is possible in a satisfactory way is as follows: Suppose that bakery $x = x_1$ is able to produce one more croissant than it already does and that café $\tilde{y}$ is short of one croissant. It might not be possible to transport the additional croissant itself to the café in need, as the costs $c(x, \tilde{y})$ may be infinite. Nevertheless it might be possible to find another bakery x_2 (which usually supplies café y_2) such that bakery x can transport (with finite costs!) the extra croissant to y_2; this leaves us with a now unused item from bakery x_2, which can be transported to $\tilde{y}$ with finite costs. Of course we allow not only one, but finitely many intermediate pairs $(x_1, y_1), \ldots, (x_n, y_n)$ of bakeries/cafés to achieve this relocation of the additional croissant.

In the Ambrosio-Pratelli example (Example 1.2) we can reroute from a point $(x, x \oplus \alpha) \in \Gamma_1$ to a point $(\tilde{x}, \tilde{x} \oplus \alpha) \in \Gamma_1$ only if there exists $n \in \mathbb{N}$ such that $x \oplus (n\alpha) = \tilde{x}$. In particular, irrationality of α implies that if we can redirect with finite costs from $(x, x \oplus \alpha)$ to $(\tilde{x}, \tilde{x} \oplus \alpha)$ we never can redirect back from $(\tilde{x}, \tilde{x} \oplus \alpha)$ to $(x, x \oplus \alpha)$. These heuristics motivate the next definition which is in the spirit of the discussion in [14, bibliographical notes to chapter 5].

Definition 1.6 *Let X, Y be Polish spaces equipped with Borel probability measures μ, ν and let $c : X \times Y \to [0, \infty]$ be a Borel measurable cost function. A c-monotone Borel set $\Gamma \subseteq X \times Y$ is* transitively c-monotone *if it has the following property: For all pairs $(x, y), (\tilde{x}, \tilde{y}) \in \Gamma$ there exist $(x_1, y_1), \ldots, (x_n, y_n) \in \Gamma$ such that $(x, y) = (x_1, y_1)$ and $(\tilde{x}, \tilde{y}) = (x_n, y_n)$ and $c(x_1, y_2), \ldots, c(x_{n-1}, y_n) < \infty$*

A transport plan π is transitively c-monotone *if there exists a transitively c-monotone set Γ with $\pi(\Gamma) = 1$.*

Every transitively c-monotone transport plan π is in particular c-monotone. More importantly, we will see that each such π is even strongly c-monotone. In fact these properties are equivalent apart from a certain subtlety:

Note that the definition of transitive c-monotonicity is sensitive to changes of the cost function $c : X \times Y \to [0, \infty]$ even on singletons, while c-monotonicity, strong c-monotonicity, optimality and robust optimality are stable with respect to changes of the cost function on "small" sets. Since we want to relate the strong and transitive notions of c-monotonicity, we have to take this disparity into account.

In the light of the following lemma the right concept of "smallness" is the notion of L-shaped null sets. Let us call a Borel set $N \subseteq X \times Y$ an *L-shaped null set* if there exist a μ-null set N_1 and a ν-null set N_2 such that $N \subseteq$

$N_1 \times Y \cup X \times N_2$.

Lemma 1.7 *Let X, Y be Polish spaces equipped with Borel probability measures μ, ν. For Borel measurable cost functions $c, \tilde{c} : X \times Y \to [0, \infty]$ the following are equivalent:*

a. $\{c \neq \tilde{c}\}$ is an L-shaped null set.
b. $I_c[\pi] = I_{\tilde{c}}[\pi]$ for every measure π on $X \times Y$ which has absolutely continuous marginals with respect to μ and ν.

Furthermore, a. and b. imply:

c. The (strongly) c-monotone transport plans coincide with the (strongly) $\tilde{c}$-monotone transport plans.
d. $I_c[\pi] = I_{\tilde{c}}[\pi]$ for all transport plans $\pi \in \Pi(\mu, \nu)$.

To achieve the desired equivalence, we have to make the harmless assumption that X (resp. Y) contains non-empty null sets with respect to μ (resp. ν).

Theorem 3 *Let X, Y be Polish spaces equipped with Borel probability measures μ, ν and $c : X \times Y \to [0, \infty]$ a Borel measurable cost function. Then a. implies b. and, if X and Y contain nonempty null sets, b. implies a.:*

a. There exist a Borel measurable function $\tilde{c} : X \times Y \to [0, \infty]$ which differs from c only on an L-shaped null set, such that π is transitively $\tilde{c}$-monotone.
b. π is strongly c-monotone.

1.4 Putting things together

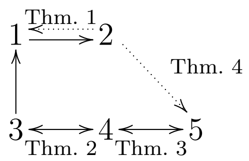
Finally we want to point out that in the situation where c is finite all previously mentioned notions of monotonicity and optimality coincide. We can even pass to a slightly more general setting than finite cost functions and obtain the following result.

Theorem 4 *Let X, Y be Polish spaces equipped with Borel probability measures μ, ν and let $c : X \times Y \to [0, \infty]$ be Borel measurable and $\mu \otimes \nu$-a.e. finite. For a finite transport plan π the following assertions are equivalent:*

(1) π is optimal.
(2) π is c-monotone.
(3) π is robustly optimal.
(4) π is strongly c-monotone.
(5) π is transitively c-monotone.

The equivalence of a., b. and d. was established in [12] under the additional assumption that c is l.s.c. and finitely valued.

We sum up the situation under fully general assumptions. The upper line (1 and 2) relates to the optimality of a transport plan π. The lower line (3), (4) and (5) of essentially equivalent conditions implies the upper line but - without additional assumptions - not vice versa.



(1) optimal

(2) c-monotone

(3) robustly optimal

(4) strongly c-monotone

(5) transitively c-monotone

Fig. 1. Implications between properties of transport plans

Note that the implications symbolized by dotted lines in Figure 1 are not true without additional assumptions. ((2) $\not\Rightarrow$ (1): Example 1.2, (1) $\not\Rightarrow$ (3) and (2) $\not\Rightarrow$ (5): Example 5.3, (4) $\not\Rightarrow$ (5): Examples 5.1 and 5.2)

The rest of the paper is organized as follows:

Section 2: Optimal $\Rightarrow$ c-monotone. (Theorem 1.a.)

Section 3: Transitively c-monotone $\Rightarrow$ strongly c-monotone.

Section 4: c-monotone $\Rightarrow$ optimal. (Theorem 1.b.)

In Section 5 we establish the remaining non-trivial implications of Figure 1. Finally, we observe that in all the above discussion we only referred to the Borel structure of the Polish spaces X, Y, and never referred to the topological structure. Hence the above results (with the exception of Theorem 1.b.) hold true for standard Borel measure spaces.

In fact it seems likely that our results can be transferred to the setting of perfect measure spaces. (See [9] for a general overview resp. [8] for a treatment of problems of mass transport in this framework.) However we do not pursue this direction.

2 Improving Transports

Assume that some transport plan $\pi \in \Pi(\mu, \nu)$ is given. From a purely heuristic point of view either the set of points on which c-monotonicity is violated is "negligible" such that π is c-monotone or there are "many" such points in which case the transport plan can be enhanced. The notion of c-monotonicity refers to n-tuples $((x_1, y_1), \dots, (x_n, y_n))$ of arbitrary length n, and it turns out that it is necessary to consider finitely many measure spaces to properly formulate what is meant by "negligible" resp. "many".

Let $X_1, \dots, X_n$ be Polish spaces equipped with finite Borel measures $\mu_1, \dots, \mu_n$. By $\Pi(\mu_1, \dots, \mu_n) \subseteq \mathcal{M}(X_1 \times \dots \times X_n)$ we denote the set of all Borel measures on $X_1 \times \dots \times X_n$ such that the i-th marginal measure coincides with the Borel measure μ_i for $i = 1, \dots, n$. By $p_{X_i} : X_1 \times \dots \times X_n \to X_i$ we denote the projection onto the i-th component. $B \subseteq X_1 \times \dots \times X_n$ is called an *L-shaped* null set if there exist null sets $N_1 \subseteq X_1, \dots, N_n \subseteq X_n$ such that $B \subseteq \bigcup_{i=1}^n p_{X_i}^{-1}[N_i]$.

The Borel sets of $X_1 \times \dots \times X_n$ satisfy a nice dichotomy. They are either L-shaped null sets or they carry a positive measure whose marginals are absolutely continuous with respect to $\mu_1, \dots, \mu_n$:

Proposition 2.1 *Let $X_1, \dots, X_n, n \geq 2$ be Polish spaces equipped with Borel probability measures $\mu_1, \dots, \mu_n$. Then for any Borel set $B \subseteq X_1 \times \dots \times X_n$ let*

$$P(B) := \sup \{\pi(B) : \pi \in \Pi(\mu_1, \dots, \mu_n)\} \tag{5}$$

$$L(B) := \inf \left\{ \sum_{i=1}^n \mu_i(B_i) : B_i \subseteq X_i \text{ and } B \subseteq \bigcup_{i=1}^n p_{X_i}^{-1}[B_i] \right\}. \tag{6}$$

Then $P(B) \geq 1/n\, L(B)$. In particular B satisfies one of the following alternatives:

a. B is an L-shaped null set.
b. There exists $\pi \in \Pi(\mu_1, \dots, \mu_n)$ such that $\pi(B) > 0$.

The main ingredient in the proof Proposition 2.1 is the following duality theorem due to Kellerer (see [5, Lemma 1.8(a), Corollary 2.18]).

Theorem (Kellerer) *Let $X_1, \dots, X_n, n \geq 2$ be Polish spaces equipped with Borel probability measures $\mu_1, \dots \mu_n$ and assume that $c : X = X_1 \times \dots \times X_n \to$*

$\mathbb{R}$ *is Borel measurable and that* $\bar{c} := \sup_X c, \underline{c} := \inf_X c$ *are finite. Set*

$$I(c) = \inf\left\{\int_X c\ d\pi : \pi \in \Pi(\mu_1, \dots, \mu_n)\right\}, \tag{7}$$

$$S(c) = \sup\left\{\sum_{i=1}^n \int_{X_i} \varphi_i\ d\mu_i : c(x_1, \dots, x_n) \le \sum_{i=1}^n \varphi_i(x_i), \frac{1}{n}\bar{c} - (\bar{c} - \underline{c}) \le \varphi_i \le \frac{1}{n}\bar{c}\right\}. \tag{8}$$

Then $I(c) = S(c)$*.*

PROOF of Proposition 2.1. Observe that $-I(-\mathbb{1}_B) = P(B)$ and that

$$-S(-\mathbb{1}_B) = \inf\left\{\sum_{i=1}^n \int_{X_i} \chi_i d\mu_i : \mathbb{1}_B(x_1, \dots, x_n) \le \sum_{i=1}^n \chi_i(x_i), 0 \le \chi_i \le 1\right\}. \tag{9}$$

By Kellerer's Theorem $-S(-\mathbb{1}_B) = -I(-\mathbb{1}_B)$. Thus it remains to show that $-S(-\mathbb{1}_B) \ge 1/n\, L(B)$. Fix functions $\chi_1, \dots, \chi_n$ as in (9). Then for each $(x_1, \dots, x_n) \in B$ one has $1 = \mathbb{1}_B(x_1, \dots, x_n) \le \sum_{i=1}^n \chi_i(x_i)$ and hence there exists some i such that $\chi_i(x_i) \ge 1/n$. Thus $B \subseteq \bigcup_{i=1}^n \pi_i^{-1}[\{\chi_i \ge 1/n\}]$. It follows that

$$\begin{aligned} -S(-\mathbb{1}_B) &\ge \inf\left\{\sum_{i=1}^n \int_{X_i} \chi_i d\mu_i : B \subseteq \bigcup_{i=1}^n p_{X_i}^{-1}[\{\chi_i \ge 1/n\}], 0 \le \chi_i \le 1\right\} \\ &\ge \inf\left\{\sum_{i=1}^n \frac{1}{n}\mu_i(\{\chi_i \ge 1/n\}) : B \subseteq \bigcup_{i=1}^n p_{X_i}^{-1}[\{\chi_i \ge 1/n\}]\right\} \ge \frac{1}{n} L(B) \end{aligned}$$

From this we deduce that either $L(B) = 0$ or that there exists $\pi \in \Pi(\mu_1, \dots, \mu_n)$ such that $\pi(B) > 0$. The last assertion of Proposition 2.1 now follows from the following Lemma due to Richárd Balka and Márton Elekes (private communication). □

Lemma 2.2 *Suppose that* $L(B) = 0$ *for a Borel set* $B \subseteq X_1 \times \dots \times X_n$*. Then* B *is an L-shaped null set.*

PROOF. Fix $\varepsilon > 0$ and Borel sets $B_1^{(k)}, \dots, B_n^{(k)}$ with $\mu_i(B_i^{(k)}) \le \varepsilon\, 2^{-k}$ such that for each k

$$B \subseteq p_{X_1}^{-1}[B_1^{(k)}] \cup \dots \cup p_{X_n}^{-1}[B_n^{(k)}].$$

Let $B_i := \bigcup_{k=1}^\infty B_i^{(k)}$ for $i = 2, \dots, n$ such that

$$B \subseteq p_{X_1}^{-1}[B_1^{(k)}] \cup p_{X_2}^{-1}[B_2] \cup \dots \cup p_{X_n}^{-1}[B_n]$$

for each $k \in \mathbb{N}$. Thus with $B_1 := \bigcap_{k=1}^\infty B_1^{(k)}$,

$$B \subseteq p_{X_1}^{-1}[B_1] \cup p_{X_2}^{-1}[B_2] \cup \dots \cup p_{X_n}^{-1}[B_n].$$

Hence we can assume from now on that $\mu_1(B_1) = 0$ and that $\mu_i(B_i)$ is arbitrarily small for $i = 2, \dots, n$. Iterating this argument in the obvious way we get the statement. $\square$

Remark 2.3 *In the case $n = 2$ it was shown in [5, Proposition 3.3] that $L(B) = P(B)$ for every Borel set $B \subseteq X_1 \times X_2$. However, for $n > 2$, equality does not hold true, cf. [5, Example 3.4].*

Proposition 2.1 allows us to prove **Lemma 1.7**.

Let X, Y be Polish spaces equipped with Borel probability measures μ, ν. For Borel measurable cost functions $c, \tilde{c} : X \times Y \to [0, \infty]$ the following are equivalent:

- *a. $\{c \neq \tilde{c}\}$ is an L-shaped null set.*
- *b. $I_c[\pi] = I_{\tilde{c}}[\pi]$ for every measure π on $X \times Y$ which has absolutely continuous marginals with respect to μ and ν.*

Furthermore, a. and b. imply:

- *c. The (strongly) c-monotone transport plans coincide with the (strongly) $\tilde{c}$-monotone transport plans.*
- *d. $I_c[\pi] = I_{\tilde{c}}[\pi]$ for all transport plans $\pi \in \Pi(\mu, \nu)$.*

PROOF. a. $\Rightarrow$ b.: It follows directly from the marginal conditions for transport plans that $\pi(N) = 0$ for every L-shaped null set N and every transport plan π. Thus it does not affect the value of $I_{\tilde{c}}[\pi]$ if we change c on an L-shaped null set.

b. $\Rightarrow$ a.: If a. does not hold true we may assume without loss of generality that $\{c > c'\}$ is not an L-shaped null set. Since L-shaped null sets are stable under forming countable unions there exists some $\varepsilon > 0$ such that $M = \{c > c' + \varepsilon\}$ is not an L-shaped null set. By Proposition 2.1 there exists a transport plan π with $\delta = \pi(M) > 0$. Restricting π to M yields the desired measure.

b. $\Rightarrow$ d.: Clear.

a. $\Rightarrow$ c.: Assume that Γ, φ and ψ witness the strong c-monotonicity of π. Pick μ- resp. ν-null sets N_1 and N_2 such that $c = \tilde{c}$ on the complement of the L-shaped null set $N = N_1 \times Y \cup X \times N_2$. Then $\tilde{\varphi}(x) = \varphi(x) - \infty \cdot \mathbb{1}_{N_1}(x)$ and $\tilde{\psi}(y) = \psi(y) - \infty \cdot \mathbb{1}_{N_2}(y)$ satisfy $\tilde{\varphi}(x) + \tilde{\psi}(y) \leq \tilde{c}(x, y)$ for all $(x, y) \in X \times Y$ and $\tilde{\varphi}(x) + \tilde{\psi}(y) = \tilde{c}(x, y)$ for all $(x, y) \in \tilde{\Gamma} = \Gamma \setminus \mathbb{N}$, thus π is strongly $\tilde{c}$-monotone.

In the case of usual c-monotonicity the statement is proved similarly. $\square$

Definition 2.4 *Let X, Y be Polish spaces. For a Borel measurable cost function $c : X \times Y \to [0, \infty]$ we define*

$$B_{n,\varepsilon} := \left\{ (x_i, y_i)_{i=1}^n \in (X \times Y)^n : \sum_{i=1}^n c(x_i, y_i) \geq \sum_{i=1}^n c(x_i, y_{i+1}) + \varepsilon \right\}. \quad (10)$$

The definition of the sets $B_{n,\varepsilon}$ is implicitly given in [2, Theorem 2.3]. The idea behind it is, that $(x_i, y_i)_{i=1}^n \in B_{n,\varepsilon}$ tells us that transport costs can be reduced if "x_i is transported to y_{i+1} instead of y_i" (recall the conventions $x_{n+1} = x_1$ resp. $y_{n+1} = y_1$). In what follows we make this statement precise and give a coordinate free formulation.

Denote by $\sigma, \tau : (X \times Y)^n \to (X \times Y)^n$ the shifts defined via

$$\sigma : (x_i, y_i)_{i=1}^n \mapsto (x_{i+1}, y_{i+1})_{i=1}^n \quad (11)$$
$$\tau : (x_i, y_i)_{i=1}^n \mapsto \quad (x_i, y_{i+1})_{i=1}^n. \quad (12)$$

Observe that $\sigma^n = \tau^n = \mathrm{Id}_{(X \times Y)^n}$ and that σ and τ commute. Also note that the set $B_{n,\varepsilon}$ from (10) is σ-invariant (i.e. $\sigma(B_{n,\varepsilon}) = B_{n,\varepsilon}$), but in general not τ-invariant. Denote by $p_i : (X \times Y)^n \to X \times Y$ the projection on the i-th component of the product. The projections $p_X : X \times Y \to X, (x, y) \mapsto x$ and $p_Y : X \times Y \to Y, (x, y) \mapsto y$ are defined as usual and shall present no danger of confusion.

Lemma 2.5 *Let X, Y be Polish spaces equipped with Borel probability measures μ, ν. Let π be a transport plan. Then one of the two alternatives holds:*

a. π is c-monotone,
b. there exist $n \in \mathbb{N}$, $\varepsilon > 0$ and a measure $\kappa \in \Pi(\pi, \ldots, \pi)$ such that $\kappa(B_{n,\varepsilon}) > 0$. Moreover κ can be taken to be both σ and τ invariant.

PROOF. Suppose that $B_{n,\varepsilon}$ is an L-shaped null set for all $n \in \mathbb{N}$ and every $\varepsilon > 0$. Then there are Borel sets $S^1_{n,\varepsilon}, \ldots, S^n_{n,\varepsilon} \subseteq X \times Y$ of full π-measure such that

$$\left(S^1_{n,\varepsilon} \times \ldots \times S^n_{n,\varepsilon} \right) \cap B_{n,\varepsilon} = \emptyset$$

and π is concentrated on the c-monotone set

$$S = \bigcap_{k=1}^{\infty} \bigcap_{n=1}^{\infty} \bigcap_{i=1}^{n} S^i_{n,1/k}.$$

If there exist $n \in \mathbb{N}$ and $\varepsilon > 0$ such that $B_{n,\varepsilon}$ is not an L-shaped null set, we apply Proposition 2.1 to conclude the existence of a measure $\kappa \in \Pi(\pi, \ldots, \pi)$

with $\kappa(B_{n,\varepsilon}) > 0$. To achieve the desired invariance, simply replace κ by

$$\frac{1}{n^2}\sum_{i,j=1}^{n}(\sigma^i \circ \tau^j)_{\#}\kappa \qquad \square \tag{13}$$

We are now in the position to prove the statement of **Theorem 1.a**, i.e.

Let X, Y be Polish spaces equipped with Borel probability measures μ, ν and let $c : X \times Y \to [0, \infty]$ be a Borel measurable cost function. If π is a finite optimal transport plan, then π is c-monotone.

PROOF. Suppose by contradiction that π is optimal, $I_c[\pi] < \infty$ but π is not c-monotone. Then by Lemma 2.5 there exist $n \in \mathbb{N}$, $\varepsilon > 0$ and an invariant measure $\kappa \in \Pi(\pi, \ldots, \pi)$ which gives mass $\alpha > 0$ to the Borel set $B_{n,\varepsilon} \subseteq (X \times Y)^n$. Consider now the restriction of κ to $B_{n,\varepsilon}$ defined via $\hat{\kappa}(A) := \kappa(A \cap B_{n,\varepsilon})$ for Borel sets $A \subseteq (X \times Y)^n$. $\hat{\kappa}$ is σ-invariant since both the measure κ and the Borel set $B_{n,\varepsilon}$ are σ-invariant. Denote the marginal of $\hat{\kappa}$ in the first coordinate $(X \times Y)$ of $(X \times Y)^n$ by $\hat{\pi}$. Due to σ-invariance we have

$$p_{i\#}\hat{\kappa} = p_{i\#}(\sigma_{\#}\hat{\kappa}) = (p_i \circ \sigma)_{\#}\hat{\kappa} = p_{i+1\#}\hat{\kappa},$$

i.e. all marginals coincide and we have $\hat{\kappa} \in \Pi(\hat{\pi}, \ldots, \hat{\pi})$. Furthermore, since $\hat{\kappa} \leq \kappa$, the same is true for the marginals, i.e. $\hat{\pi} \leq \pi$. Denote the marginal of $\tau_{\#}\hat{\kappa}$ in the first coordinate $(X \times Y)$ of $(X \times Y)^n$ by $\hat{\pi}_\beta$. As σ and τ commute, $\tau_{\#}\hat{\kappa}$ is σ-invariant, so the marginals in the other coordinates coincide with $\hat{\pi}_\beta$. An easy calculation shows that $\hat{\pi}$ and $\hat{\pi}_\beta$ have the same marginals in X resp. Y:

$$\begin{aligned}
p_{X\#}\hat{\pi}_\beta &= p_{X\#}(p_{i\#}(\tau_{\#}\hat{\kappa})) = (p_X \circ p_i \circ \tau)_{\#}\hat{\kappa} = (p_X \circ p_i)_{\#}\hat{\kappa} = p_{X\#}\hat{\pi}, \\
p_{Y\#}\hat{\pi}_\beta &= p_{Y\#}(p_{i\#}(\tau_{\#}\hat{\kappa})) = (p_Y \circ p_i \circ \tau)_{\#}\hat{\kappa} = (p_Y \circ p_{i+1})_{\#}\hat{\kappa} = p_{Y\#}\hat{\pi}.
\end{aligned}$$

The equality of the total masses is proved similarly:

$$\alpha := \hat{\pi}_\beta(X \times Y) = (p_i \circ \tau)_{\#}\hat{\kappa}(X \times Y) = p_{i\#}\hat{\kappa}(X \times Y) = \hat{\pi}(X \times Y).$$

Next we compute the transport costs associated to $\hat{\pi}_\beta$:

$$\begin{aligned}
\int_{X\times Y} c\,d\hat{\pi}_\beta = & \int_{(X\times Y)^n} c\circ p_1\, d(\tau_\#\hat{\kappa}) && \text{(marginal property)}\\
= & \frac{1}{n}\sum_{i=1}^n \int_{(X\times Y)^n} c\circ p_i\, d(\tau_\#\hat{\kappa}) && (\sigma\text{-invariance})\\
= & \frac{1}{n}\sum_{i=1}^n \int_{(X\times Y)^n} (c\circ p_i\circ\tau)\, d\kappa && \text{(push-forward)}\\
= & \frac{1}{n}\sum_{i=1}^n \int_{B_{n,\varepsilon}} (c\circ p_i\circ\tau)\, d\kappa && \text{(definition of } \hat{\kappa})\\
\leq & \frac{1}{n}\int_{B_{n,\varepsilon}} \left[\sum_{i=1}^n (c\circ p_i)\ -\varepsilon\right] d\kappa && \text{(definition of } B_{n,\varepsilon})\\
= & \int_{X\times Y} c d\hat{\pi} - \varepsilon\frac{\alpha}{n}. && \text{(definition of } \hat{\pi})
\end{aligned}$$

To improve the transport plan π we define

$$\pi_\beta := (\pi - \hat{\pi}) + \hat{\pi}_\beta. \tag{14}$$

Recall that $\pi - \hat{\pi}$ is a positive measure, so π_β is a positive measure. As $\hat{\pi}$ and $\hat{\pi}_\beta$ have the same total mass, π_β is a probability measure. Furthermore $\hat{\pi}$ and $\hat{\pi}_\beta$ have the same marginals, so π_β is indeed a transport plan. We have

$$I_c[\pi_\beta] = I_c[\pi] + \int_{X\times Y} c\, d(\hat{\pi}_\beta - \hat{\pi}) \leq I_c[\pi] - \varepsilon\frac{\alpha}{n} < I_c[\pi]. \qquad \square \tag{15}$$

3 Establishing strong c-monotonicity

The objective of this section is to prove Proposition 3.1, based on several lemmas which will be introduced throughout the section.

Proposition 3.1 *Let X, Y be Polish spaces equipped with Borel probability measures μ, ν and let $c : X \times Y \to [0,\infty]$ be a Borel measurable cost function. Then every transitively c-monotone transport plan π is strongly c-monotone.*

In the course of the proof of Proposition 3.1 we need the following definition.

Definition 3.2 *Let X, Y be Polish spaces equipped with Borel probability measures μ, ν, let $c : X \times Y \to [0,\infty]$ be a Borel measurable cost function and $\Gamma \subseteq X \times Y$ a Borel measurable set on which c is finite. We define*

a. $(x, y) \lesssim (\tilde{x}, \tilde{y})$ *if there exist pairs* $(x_0, y_0), \ldots, (x_n, y_n) \in \Gamma$ *such that* $(x, y) = (x_0, y_0)$ *and* $(\tilde{x}, \tilde{y}) = (x_n, y_n)$ *and* $c(x_1, y_0), \ldots, c(x_n, y_{n-1}) < \infty$.
b. $(x, y) \approx (\tilde{x}, \tilde{y})$ *if* $(x, y) \lesssim (\tilde{x}, \tilde{y})$ *and* $(x, y) \gtrsim (\tilde{x}, \tilde{y})$.

We call (Γ, c) connecting *if* c *is finite on* Γ *and* $(x, y) \approx (\tilde{x}, \tilde{y})$ *for all* $(x, y), (\tilde{x}, \tilde{y}) \in \Gamma$.

Thus Γ is transitively c-monotone if and only if it is c-monotone and (Γ, c) is connecting. These relations were introduced in [14, Chapter 5, p.75] and appear in a construction due to Stefano Bianchini.

When there is any danger of confusion we will write $\lesssim_{c,\Gamma}$ and $\approx_{c,\Gamma}$, indicating the dependence on Γ and c. Note that $\lesssim$ is a pre-order, i.e. a transitive and reflexive relation, and that $\approx$ is an equivalence relation. We will also need the projections $\lesssim_X, \approx_X$ resp. $\lesssim_Y, \approx_Y$ of these relations onto the set $p_X[\Gamma] \subseteq X$ resp. $p_Y[\Gamma] \subseteq Y$. The projection is defined in the obvious way: $x \lesssim_X \tilde{x}$ if there exist $y, \tilde{y}$ such that $(x, y), (\tilde{x}, \tilde{y}) \in \Gamma$ and $(x, y) \lesssim (\tilde{x}, \tilde{y})$ holds.

The other relations are defined analogously. The projections of $\lesssim$ are again pre-orders and the projections of $\approx$ are again equivalence relations, provided c is finite on Γ. The equivalence classes of $\approx$ and its projections are compatible in the sense that $[(x, y)]_\approx = ([x]_{\approx_X} \times [y]_{\approx_Y}) \cap \Gamma$. The elementary proofs of these facts are left to the reader.

In the proof of Proposition 3.1 we will establish the existence of the functions φ, ψ using the construction given in [11], see also [13, Chapter 2] and [1, Theorem 3.2]. As we do not impose any continuity assumptions on the cost function c, we can not prove the Borel measurability of φ and ψ by using limiting procedures similar to the methods used in [1,11–13]. Instead we will employ certain concepts from descriptive set theory:

Let X be a Polish space. Given a Borel measure μ on X, we denote its completion by $\tilde{\mu}$. A subset of X is called *universally measurable* if it is measurable with respect to the completion of every σ-finite Borel measure on X. The universally measurable sets of X form a σ-algebra which is strictly larger than the Borel sets. Due to a result of Luzin [4, Theorem 21.10] all *analytic* sets are universally measurable. By definition, a set $A \subseteq X$ is analytic if there exist a Polish space Y, a Borel measurable function $f : Y \to X$ and a Borel set $B \subseteq Y$ such that $f(B) = A$.

We call a function $f : X \to [-\infty, \infty]$ universally measurable if the pre-image of every Borel set is universally measurable.

Lemma 3.3 *Let* X *be a Polish space and* μ *a finite Borel measure on* X*. If* $\varphi : X \to \mathbb{R} \cup \{-\infty\}$ *is universally measurable, then there exists a Borel measurable function* $\tilde{\varphi} : X \to \mathbb{R} \cup \{-\infty\}$ *such that* $\tilde{\varphi} \leq \varphi$ *everywhere and* $\varphi = \tilde{\varphi}$ *almost everywhere with respect to* $\tilde{\mu}$*.*

PROOF. Let $(I_n)_{n=1}^{\infty}$ be an enumeration of the intervals $[a, b)$ with endpoints in $\mathbb{Q}$. Then for each $n \in \mathbb{N}$, $\varphi^{-1}[I_n]$ is $\tilde{\mu}$-measurable and hence the union of a Borel set B_n and a $\tilde{\mu}$-null set N_n. Let N be a Borel null set which covers $\bigcup_{n=1}^{\infty} N_n$. Let $\tilde{\varphi}(x) = \varphi(x) - \infty \cdot \mathbf{1}_N(x)$. Clearly $\tilde{\varphi}(x) \leq \varphi(x)$ for all $x \in X$ and $\varphi(x) = \tilde{\varphi}(x)$ for $\tilde{\mu}$-almost all $x \in X$. Furthermore, $\tilde{\varphi}$ is Borel measurable since $(I_n)_{n=1}^{\infty}$ is a generator of the Borel σ-algebra on $\mathbb{R} \cup \{-\infty\}$ and for each $n \in \mathbb{N}$ we have that $\tilde{\varphi}^{-1}[I_n] = B_n \setminus N$ is a Borel set. □

The following definition of the functions $\varphi_n, n \in \mathbb{N}$ resp. φ is reminiscent of the construction in [11].

Lemma 3.4 *Let X, Y be Polish spaces, $c : X \times Y \to [0, \infty]$ a Borel measurable cost function and $\Gamma \subseteq X \times Y$ a Borel set. Fix $(x_0, y_0) \in \Gamma$ and assume that c is finite on Γ. For $n \in \mathbb{N}$, define $\varphi_n : X \times \Gamma^n \to \mathbb{R} \cup \{\infty\}$ by*

$$\varphi_n(x; x_1, y_1, \ldots, x_n, y_n) = [c(x, y_n) - c(x_n, y_n)] + \sum_{i=0}^{n-1} [c(x_{i+1}, y_i) - c(x_i, y_i)] \quad (16)$$

Then the map $\varphi : X \to \mathbb{R} \cup \{\pm\infty\}$ defined by

$$\varphi(x) = \inf\{\varphi_n(x; x_1, y_1, \ldots, x_n, y_n) : n \geq 1, (x_i, y_i)_{i=1}^n \in \Gamma^n\} \quad (17)$$

is universally measurable.

PROOF. First note that the Borel σ-algebra on $\mathbb{R} \cup \{\pm\infty\}$ is generated by intervals of the form $[-\infty, \alpha)$, thus it is sufficient to determine the pre-images of those sets under φ. We have

$$\varphi(x) < \alpha \leftrightarrow \exists n \in \omega\ \exists (x_0, y_0), \ldots, (x_n, y_n) \in \Gamma : \varphi_n(x; x_0, y_0, \ldots, x_n, y_n) < \alpha.$$

The set $\varphi_n^{-1}[[-\infty, \alpha)]$ is Borel measurable. Hence

$$\varphi^{-1}[[-\infty, \alpha)] = \bigcup_{n \in \omega} p_X[\varphi_n^{-1}[[-\infty, \alpha)]]$$

is the countable union of projections of Borel sets. Since projections of Borel sets are analytic, $\varphi^{-1}[[-\infty, \alpha)]$ belongs to the σ-algebra of universally measurable sets. □

Lemma 3.5 *Let X, Y be Polish spaces and $c : X \times Y \to [0, \infty]$ a Borel measurable cost function. Suppose Γ is c-monotone, c is finite on Γ and (Γ, c) is connecting. Fix $(x_0, y_0) \in \Gamma$. Then the map φ from (17) is finite on $p_X[\Gamma]$. Furthermore*

$$\varphi(x) \leq \varphi(x') + c(x, y) - c(x', y) \quad \forall x \in X,\ (x', y) \in \Gamma. \quad (18)$$

PROOF. Fix $x \in p_X[\Gamma]$. Since $x_0 \lesssim x$ (recall Definition 3.2), we can find $x_1, y_1, \ldots, x_n, y_n$ such that $\varphi_n(x; x_1, y_1, \ldots, x_n, y_n) < \infty$. Hence $\varphi(x) < \infty$. Proving $\varphi(x) > -\infty$ involves some wrestling with notation but, not very surprisingly, it comes down to applying the fact that $x \lesssim x_0$. Let $a_1 = x$ and choose $b_1, a_2, b_2, \ldots, a_m, b_m$ such that $(a_1, b_1), \ldots, (a_m, b_m) \in \Gamma$ and $c(a_2, b_1), \ldots, c(a_m, b_{m-1}), c(x, b_m) < \infty$. Assume now that $x_1, y_1, \ldots, x_n, y_n$ are given such that $\varphi_n(x; x_1, y_1, \ldots, x_n, y_n) < \infty$. Put $x_{n+i} = a_i$ and $y_{n+i} = b_i$ for $i \in \{1, \ldots, m\}$. Due to c-monotonicity of Γ and the finiteness of all involved terms we have:

$$0 \leq [c(x_0, y_{n+m}) - c(x_{n+m}, y_{n+m})] + \sum_{i=0}^{n+m-1} [c(x_{i+1}, y_i) - c(x_i, y_i)],$$

which, after regrouping yields

$$\begin{aligned} \alpha := [c(x_0, b_m) - c(a_m, b_m)] + \sum_{i=1}^{m-1} [c(a_{i+1}, b_i) - c(a_i, b_i)] \\ \leq [c(x, y_n) - c(x_n, y_n)] + \sum_{i=0}^{n-1} [c(x_{i+1}, y_i) - c(x_i, y_i)]. \end{aligned} \tag{19}$$

Note that the right hand side of (19) is just $\varphi_n(x; x_1, y_1, \ldots, x_n, y_n)$. Thus passing to the infimum we see that $\varphi(x) \geq \alpha > -\infty$. To prove the remaining inequality, observe that the right hand side of (18) can be written as

$$\inf\{\varphi_n(x; x_1, y_1, \ldots, x_n, y_n) : n \geq 1, (x_i, y_i)_{i=1}^n \in \Gamma^n \text{ and } (x_n, y_n) = (x', y)\}$$

whereas the left hand side of (18) is the same, without the restriction $(x_n, y_n) = (x', y)$. $\square$

Lemma 3.6 *Let X, Y be Polish spaces and $c : X \times Y \to [0, \infty]$ a Borel measurable cost function. Let $X_0 \subseteq X$ be a nonempty Borel set and let $\varphi : X_0 \to \mathbb{R}$ be a Borel measurable function. Then the c-transform $\psi : Y \to \mathbb{R} \cup \{-\infty\}$, defined as*

$$\psi(y) := \inf_{x \in X_0} [c(x, y) - \varphi(x)] \tag{20}$$

is universally measurable.

PROOF. As in the proof of Lemma 3.3 we consider the set $\psi^{-1}[[-\infty, \alpha)]$:

$$\psi(y) < \alpha \leftrightarrow \exists x \in X_0 : c(x, y) - \varphi(x) < \alpha.$$

Note that the set $\{(x, y) \in X_0 \times Y : c(x, y) - \varphi(x) < \alpha\}$ is Borel. Thus

$$\psi^{-1}[[-\infty, \alpha)] = p_X[\{(x, y) \in X_0 \times Y : c(x, y) - \varphi(x) < \alpha\}]$$

is the projection of a Borel set, i.e. analytic and hence universally measurable. $\square$

We are now able to prove **Proposition 3.1**.

Let X, Y be Polish spaces equipped with Borel probability measures μ, ν and let $c : X \times Y \to [0, \infty]$ be a Borel measurable cost function. Then every transitively c-monotone transport plan π is strongly c-monotone.

PROOF. Let $\Gamma \subseteq X \times Y$ be a c-monotone Borel set such that $\pi(\Gamma) = 1$ and the pair (Γ, c) is connecting. Let φ be the map from Lemma 3.4. Using Lemma 3.3 and Lemma 3.5, and eventually passing to a subset of full π-measure, we may assume that φ is Borel measurable, that $X_0 := p_X[\Gamma]$ is a Borel set and that

$$c(x', y) - \varphi(x') \leq c(x, y) - \varphi(x) \quad \forall x \in X_0,\ (x', y) \in \Gamma. \tag{21}$$

Note that (21) follows from (18) in Lemma 3.5. Here we consider $x \in X_0$ in order to ensure that $\varphi(x)$ is finite on X_0. Now consider the c-transform

$$\psi(y) := \inf_{x \in X_0} [c(x, y) - \varphi(x)], \tag{22}$$

which by Lemma 3.6 is universally measurable. Fix $y \in p_Y[\Gamma]$. Using (21) we see that the infimum in (22) is attained at a point $x_0 \in X_0$ satisfying $(x_0, y) \in \Gamma$. This implies that $\varphi(x) + \psi(y) = c(x, y)$ on Γ and $\varphi(x) + \psi(y) \leq c(x, y)$ on $p_X[\Gamma] \times p_Y[\Gamma]$. To guarantee this inequality on the whole product $X \times Y$, one has to redefine φ and ψ to be $-\infty$ on the complement of $p_X[\Gamma]$ resp. $p_Y[\Gamma]$. Applying Lemma 3.3 once more, we find that there exists a Borel set $N \subseteq Y$ of zero ν-measure, such that $\tilde{\psi}(y) = \psi(y) - \infty \cdot \mathbb{1}_N(y)$ is Borel measurable. Finally, replace Γ by $\Gamma \cap (X \times (Y \setminus N))$ and ψ by $\tilde{\psi}$. $\square$

Finally by Proposition 3.1 the implication a. $\Rightarrow$ b. in **Theorem 3** holds true.

Let X, Y be Polish spaces equipped with Borel probability measures μ, ν and $c : X \times Y \to [0, \infty]$ a Borel measurable cost function. For a finite transport plan π, a. implies b.:

- *a. There exists a Borel measurable function $\tilde{c} : X \times Y \to [0, \infty]$, which differs from c only on an L-shaped null set, such that π is transitively $\tilde{c}$-monotone.*
- *b. π is strongly c-monotone.*

4 From c-monotonicity to optimality

This section is devoted to the proof Theorem 1.b. Our argument starts with a finite c-monotone transport plan π and we aim for showing that π is at least as good as any other finite transport plan. The idea behind the proof is to partition X and Y into cells $C_i, i \in I$ resp. $D_i, i \in I$ in such a way that π is strongly c-monotone on "diagonal" sets of the form $C_i \times D_i$ while regions $C_i \times D_j, i \neq j$ can be ignored, because no finite transport plan will give positive measure to the set $C_i \times D_j$.

Thus it will be necessary to apply previously established results to some restricted transport problems on a space $C_i \times D_i$ equipped with some relativized transport plan $\pi \upharpoonright C_i \times D_i$. As in general the cells C_i, D_i are plainly Borel sets they may fail to be Polish spaces with respect to the topologies inherited from X resp. Y. However, for us it is only important that there exist *some* Polish topologies that generate the same Borel sets on C_i resp. D_i (see e.g. [4, Section 08.15]). At this point it is crucial that our results only need measurability of the cost function and do not ask for any form of continuity (cf. the remarks at the end of the introduction). Before we a give the proof of Theorem 1.b we will need some preliminary lemmas.

Lemma 4.1 *Let X, Y be Polish spaces equipped with Borel probability measures μ, ν and let $c : X \times Y \to [0, \infty]$ be a Borel measurable cost function. Let π, π_0 be finite transport plans and $\Gamma \subseteq X \times Y$ a Borel set with $\pi(\Gamma) = 1$. Let $I = \{0, \ldots, n\}$ or $I = \mathbb{N}$ and assume that $C_i, i \in I$ are mutually disjoint Borel sets in X, $D_i, i \in I$ are mutually disjoint Borel sets in Y such that the equivalence classes of $\approx_{c,\Gamma}$ are of the form $\Gamma \cap (C_i \times D_i)$. Then also $\pi(\bigcup_{i \in I} C_i \times D_i) = 1$.*

We will use some facts on Markov chains with at most countable state space I. Denote by p_{ij} the probability of transition from state i to state j. A state i is *recurrent* if one returns to the state i almost surely. A state i is called *absorbing* if one almost surely stays at state i, i.e. if $p_{ii} = 1$.

PROOF of Lemma 4.1 As $\approx_{\Gamma,c}$ is an equivalence relation and π is concentrated on Γ the sets C_i, $i \in I$ are a partition of X modulo μ-null sets. Likewise the sets D_i, $i \in I$ form a partition of Y modulo ν-null sets. In particular the quantities

$$p_i := \mu(C_i) = \nu(D_i) = \pi(C_i \times D_i), \quad i \in I \tag{23}$$

add up to 1. Without loss of generality we may assume that $p_i > 0$ for all $i \in I$. We define

$$p_{ij} := \frac{\pi_0(C_i \times D_j)}{\mu(C_i)}, \quad i, j \in I. \tag{24}$$

Then $P = (p_{ij})_{i,j \in I}$ is a stochastic matrix and $p = (p_i)_{i \in I}$ is a stochastic vector. By the condition on the marginals of π_0 we have for the i-th component of $p \cdot P$

$$(p \cdot P)_i = \sum_{j \in I} \mu(C_j) \frac{\pi_0(C_j \times D_i)}{\mu(C_j)} = \nu(D_i) = p_i$$

i.e. $p \cdot P = p$. In terms of the associated Markov chain with state space I, this means that p is an invariant distribution. By [3, Proposition 8.13] a state i of an invariant distribution $(p_i)_{i \in I}$ is recurrent if $p_i > 0$ such that in our case all states are recurrent.

We show that every state $i \in I$ is even absorbing: Suppose for contradiction that there exists some state $i = i_0$ which is not absorbing, that is $p_{ii} < 1$. Then there exists some state $i_1 \neq i$ such that $p_{ii_1} > 0$. Since i is recurrent it is in particular possible to reach i_0 starting from i_1, i.e. there exit states $i_2, \ldots, i_n$ where $i_n = i_0$ such that $p_{i_1 i_2}, \ldots, p_{i_{n-1} i_n} > 0$. Fix $k \in \{1, \ldots, n-1\}$. Then

$$\pi_0(C_{i_k} \times D_{i_{k+1}}) = p_{i_k i_{k+1}} > 0.$$

Since π_0 is a finite transport plan, there exist $x_k \in C_{i_k} \cap p_X[\Gamma]$ and $y'_{k+1} \in D_{i_{k+1}} \cap p_Y[\Gamma]$ such that $c(x_k, y'_{k+1}) < \infty$. Choose $y_k \in D_{i_k}$ and $x'_{k+1} \in C_{i_{k+1}}$ such that $(x_k, y_k), (x'_{k+1}, y'_{k+1}) \in \Gamma$. Then

$$(x_0, y_0) \lesssim (x'_1, y'_1) \approx (x_1, y_1) \lesssim (x'_2, y'_2) \approx (x_2, y_2) \lesssim \ldots \lesssim (x'_n, y'_n) \approx (x_0, y_0). \tag{25}$$

But this implies that $(x_0, y_0) \approx (x_1, y_1)$, contradicting the assumption that $(C_{i_0} \times D_{i_0}) \cap \Gamma, (C_{i_1} \times D_{i_1}) \cap \Gamma$ are different equivalence classes of $\approx_{\Gamma,c}$. Hence every state $i \in I$ is in fact absorbing, thus for all $i \in I, \pi_0(C_i \times D_i) = \mu(C_i)$ which implies $\pi_0(\bigcup_{i \in I} C_i \times D_i) = 1$. □

Lemma 4.2 *Let X, Y be Polish spaces equipped with Borel probability measures μ, ν and let $c : X \times Y \to [0, \infty]$ be a Borel measurable cost function which is $\mu \otimes \nu$-a.e. finite. For every finite transport plan π and every Borel set $\Gamma \subseteq X \times Y$ with $\pi(\Gamma) = 1$ there exist Borel sets $O \subseteq X, U \subseteq Y$ such that $\Gamma' = \Gamma \cap (O \times U)$ has full π-measure and (Γ', c) is connecting. In particular, every c-monotone transport plan is transitively c-monotone, provided that c is $\mu \otimes \nu$-finite.*

PROOF. Assume without loss of generality that c is finite on Γ. By Fubini's Theorem for μ-almost all $x \in X$ the set $\{y : c(x, y) < \infty\}$ has full ν-measure and for ν-almost all $y \in Y$ the set $\{x : c(x, y) < \infty\}$ has full μ-measure. In particular the set of points (x_0, y_0) such that both $\mu(\{x : c(x, y_0) < \infty\}) = 1$ and $\nu(\{y : c(x_0, y) < \infty\}) = 1$ has full π-measure. Fix such a pair $(x_0, y_0) \in \Gamma$ and let $O = \{x \in X : c(x, y_0) < \infty\}, U = \{y \in Y : c(x_0, y) < \infty\}$. Then $\Gamma' = \Gamma \cap (O \times U)$ has full π-measure and for every $(x, y) \in \Gamma'$ both quantities

$c(x, y_0)$ and $c(x_0, y)$ are finite. Hence $x \approx_X x_0$, for every $x \in p_X[\Gamma']$. Similarly we obtain $y \approx_Y y_0$, for every $y \in p_Y[\Gamma']$. Hence (Γ', c) is connecting. $\square$

Finally we prove the statement of **Theorem 1.b**:

Let X, Y be Polish spaces equipped with Borel probability measures μ, ν and $c : X \times Y \to [0, \infty]$ a Borel measurable cost function. Every finite c-monotone transport plan is optimal if there exist a closed set F and a $\mu \otimes \nu$-null set N such that $\{(x, y) : c(x, y) = \infty\} = F \cup N$.

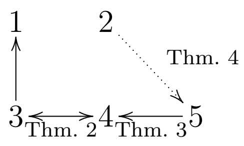
PROOF. Let c' be a function which differs from c only on N and satisfies $\{(x, y) : c'(x, y) = \infty\} = F$. Choose open sets $O_n, U_n, n \in \mathbb{N}$ such that Γ is contained in $\bigcup_{n \in \mathbb{N}}(O_n \times U_n)$ and c' is finite on this set. Fix $n \in \mathbb{N}$ and interpret $\pi \upharpoonright O_n \times U_n$ as a transport plan on the spaces (O_n, μ_n) and (U_n, ν_n) where μ_n and ν_n are the marginals corresponding to $\pi \upharpoonright O_n \times U_n$. Apply Lemma 4.2 to $\Gamma \cap (O_n \times U_n)$ and the cost function $c \upharpoonright X \times Y$ to find $O'_n \subseteq O_n, U'_n \subseteq U_n$ and $\Gamma_n = \Gamma \cap (O'_n \times U'_n)$ with $\pi(\Gamma_n) = \pi(\Gamma \cap (O_n \times U_n))$ such that (Γ_n, c) is connecting. Then $\tilde{\Gamma} = \bigcup_{n \in \mathbb{N}} \Gamma_n$ is a subset of Γ of full measure and every equivalence class of $\approx_{\tilde{\Gamma},c}$ can be written in the form $((\bigcup_{n \in N} O_n) \times (\bigcup_{n \in N} U_n)) \cap \Gamma$ for some non-empty index set $N \subseteq \mathbb{N}$. Thus there are at most countably many equivalence classes which we can write in the form $(C_i \times D_i) \cap \Gamma, i \in I$ where $I = \{1, \ldots, n\}$ or $I = \mathbb{N}$. Note that by shrinking the sets $C_i, D_i, i \in I$ we can assume that $C_i \cap C_j = D_i \cap D_j = \emptyset$ for $i \neq j$.

Assume now that we are given another finite transport plan π_0. Apply Lemma 4.1 to π, π_0 and $\tilde{\Gamma}$ to achieve that $\tilde{\pi}$ is concentrated on $\bigcup_{i \in I} C_i \times D_i$. For $i \in I$ we consider the restricted problem of transporting $\mu \upharpoonright C_i$ to $\nu \upharpoonright D_i$. We know that $\pi \upharpoonright C_i \times D_i$ is optimal for this task by Proposition 3.1, hence $I_c[\pi] \leq I_c[\pi_0]$. $\square$

Remark 4.3 *In fact the following somewhat more general (but also more complicated to state) result holds true:* Assume that $\{(x, y) : c(x, y) = \infty\} \subseteq F \cup N$ where F is closed and N is a $\mu \otimes \nu$-null set. Then every c-monotone transport plan π with $\pi(F \cup N) = 0$ is optimal.

5 Completing the picture

The most involved statements (including Theorem 1) of this paper have been proved in the previous sections. It remains to put the pieces together to establish the remaining results mentioned in the introduction (in particular Theorems 2, 3 and 4). For the convenience of the reader we collect the missing implications in Figure 5 below.



(1) optimal

(2) c-monotone

(3) robustly optimal

(4) strongly c-monotone

(5) transitively c-monotone

Fig. 2. Implications between properties of transport plans

Clearly, robust optimality (3) implies optimality (1). The first statement which we will prove is **Theorem 2**.

Let X, Y be Polish spaces equipped with Borel probability measures μ, ν and $c : X \times Y \to [0, \infty]$ a Borel measurable cost function. For a finite transport plan π the following assertions are equivalent:

(3) π is robustly optimal.
(4) π is strongly c-monotone.

PROOF. (3) $\Rightarrow$ (4): Let Z and λ be according to the definition of robust optimality. As $\tilde{\pi} = (\mathrm{Id}_Z \times \mathrm{Id}_Z)_{\#}\lambda + \pi$ is optimal, Theorem 1.a ensures the existence of a $\tilde{c}$-monotone Borel set $\tilde{\Gamma} \subseteq (X \cup Z) \times (Y \cup Z)$ such that $\tilde{c}$ is finite on $\tilde{\Gamma}$ and $\tilde{\pi}$ is concentrated on $\tilde{\Gamma}$. Note that $(z, z) \in \tilde{\Gamma}$ for λ-a.e. $z \in Z$. We claim that for λ-a.e. $z \in Z$ and all $(x, y) \in \Gamma = \tilde{\Gamma} \cap (X \times Y)$ the relation

$$(x, y) \approx_{\tilde{\Gamma}, \tilde{c}} (z, z)$$

holds true. Due to finiteness of $\tilde{c}$ on $(X \cup Z) \times Z$ respectively on $Z \times (Y \cup Z)$ we see this directly by considering the path joining (x, y) with (z, z). Moreover $(x, y) \approx_{\tilde{\Gamma}, \tilde{c}} (\tilde{x}, \tilde{y})$ for all $(x, y), (\tilde{x}, \tilde{y}) \in \tilde{\Gamma} \setminus X \times Y$, since $\tilde{c}$ is finite on the complement of $X \times Y$. Thus $(\tilde{\Gamma}, \tilde{c})$ is connecting. Applying Proposition 3.1 to the spaces $X \cup Z$ and $Y \cup Z$ we get that $\tilde{\pi}$ is strongly $\tilde{c}$-monotone, i.e. there exist $\tilde{\varphi}$ and $\tilde{\psi}$ such that $\tilde{\varphi}(a) + \tilde{\psi}(b) \leq \tilde{c}(a, b)$ and equality holds $\tilde{\pi}$-almost everywhere. By restricting $\tilde{\varphi}$ and $\tilde{\psi}$ to X resp. Y we see that π is strongly c-monotone.

(4) $\Rightarrow$ (3): Let Z be a Polish space and let λ be a positive finite Borel measure on Z. We extend c to $\tilde{c} : (X \cup Z) \times (Y \cup Z) \to [0, \infty]$ via

$$\tilde{c}(a,b) = \begin{cases} c(a,b) & \text{for} \quad (a,b) \in X \times Y \\ \max(\varphi(a), 0) & \text{for} \quad (a,b) \in X \times Z \\ \max(\psi(b), 0) & \text{for} \quad (a,b) \in Z \times Y \\ 0 & \text{otherwise.} \end{cases}$$

Define $\tilde{\varphi}(x) := \begin{cases} \varphi(a) & \text{for} \quad a \in X \\ 0 & \text{for} \quad a \in Z \end{cases}$ and $\tilde{\psi}(x) := \begin{cases} \varphi(b) & \text{for} \quad b \in Y \\ 0 & \text{for} \quad b \in Z \end{cases}$

Then $\tilde{\varphi}$ resp. $\tilde{\psi}$ are extensions of φ resp. ψ to $X \cup Z$ resp. $Y \cup Z$ which satisfy $\tilde{\varphi}(x) + \tilde{\psi}(y) \leq \tilde{c}(x,y)$ and equality holds on $\tilde{\Gamma} = \Gamma \cup \{(z,z) : z \in Z\}$. Hence $\tilde{\Gamma}$ is strongly $\tilde{c}$-monotone. Since $\tilde{\pi}$ is concentrated on $\tilde{\Gamma}$, $\tilde{\pi}$ is optimal by the Borel measurable version of [12, Proposition 2.1]. $\Box$

We start to prove the remaining part of **Theorem 3**:

Let X, Y be Polish spaces equipped with Borel probability measures μ, ν and $c : X \times Y \to [0, \infty]$ a Borel measurable cost function. Then, if X and Y contain nonempty null sets, (5) implies (4):

(4) π is strongly c-monotone.
(5) There exist a Borel measurable function $\tilde{c} : X \times Y \to [0, \infty]$, which differs from c only on an L-shaped null set, such that π is transitively $\tilde{c}$-monotone.

PROOF. Let $\Gamma \subseteq X \times Y$ be a strongly c-monotone Borel set with $\pi(\Gamma) = 1$. As both μ and ν have nonempty null sets we can pick (x_0, y_0) such that neither x_0 is an atom of μ nor y_0 is an atom of ν. Let $\varphi : X \to \mathbb{R} \cup \{-\infty\}$, $\psi : Y \to \mathbb{R} \cup \{-\infty\}$ be Borel functions which witness the strong c-monotonicity of Γ. Define

$$\tilde{\varphi}(x) := \begin{cases} \varphi(x) & \text{for} \quad x \neq x_0 \\ 0 & \text{for} \quad x = x_0 \end{cases}, \qquad \tilde{\psi}(x) := \begin{cases} \psi(y) & \text{for} \quad y \neq y_0 \\ 0 & \text{for} \quad y = y_0 \end{cases}$$

and set $\tilde{\Gamma} = \Gamma \cup \{(x_0, y_0)\}$. Alter the function c on the L-shaped null set $\{x_0\} \times Y \cup X \times \{y_0\}$ by setting

$$\tilde{c}(x_0, y) = \max(\tilde{\psi}(y), 0),$$
$$\tilde{c}(x, y_0) = \max(\tilde{\varphi}(x), 0)$$

and $\tilde{c}(x, y) = c(x, y)$ otherwise. Thus $\tilde{c}$ satisfies $\tilde{c}(x_0, y) < \infty$ for $y \in p_Y[\Gamma]$ and $\tilde{c}(x, y_0) < \infty$ for $x \in p_X[\Gamma]$. It is then obvious that $(\tilde{\Gamma}, \tilde{c})$ is connecting, Since all points lie in the same class as (x_0, y_0). Moreover $\tilde{\Gamma}$ is $\tilde{c}$-monotone since the pair of Borel functions $(\tilde{\varphi}, \tilde{\psi})$ witnesses that $\tilde{\Gamma}$ is even strongly $\tilde{c}$-monotone. □

By the discussion following Definition 1.6 we would not expect the implication (4) ⇒ (5) to be true without allowing changes of c on an L-shaped null set (strong c-monotonicity is stable under changes of c on an L-shaped null set, but transitive c-monotonicity is not). This is witnessed by the following simple example:

Example 5.1 *Let $X = Y = [0, 1]$ equipped with Lebesgue measure $\lambda = \mu = \nu$. Let $\Gamma = \{(x, x) : x \in X\}$ be the diagonal and consider the cost function $c : X \times Y \to [0, \infty]$ defined by $c = 0$ on Γ and $c = \infty$ otherwise. Clearly there is only one measure $\pi \in \Pi(\mu, \nu)$ with finite costs and π is concentrated on Γ. The constant functions $\varphi(x) = \psi(y) = 0$ witness that Γ is strongly c-monotone. However for every $\Gamma' \subseteq \Gamma$, the $\approx_{\Gamma',c}$-equivalence classes are singletons, i.e. π is not transitively c-monotone.*

If every null set contained in X or Y is empty, the following rather trivial example shows that (5) ⇒ (4) does not hold.

Example 5.2 *Let $X = Y = \{0, 1\}$, let $\mu = \nu = (\delta_0 + \delta_1)/2$ and define $c : X \times Y \to [0, \infty]$ to be 0 on $\Gamma = \{(0, 0), (1, 1)\}$ and ∞ otherwise. Then the unique transport plan π supported by Γ is strongly c-monotone but not transitively c-monotone. Since there exist no non-trivial L-shaped null sets, there is no possibility to alter c in a way which makes π transitively c-monotone.*

Finally we will establish **Theorem 4**.

Let X, Y be Polish spaces equipped with Borel probability measures μ, ν and let $c : X \times Y \to [0, \infty]$ be Borel measurable and $\mu \otimes \nu$-a.e. finite. For a finite transport plan π the following assertions are equivalent:

(1) π is optimal.
(2) π is c-monotone.
(3) π is robustly optimal.
(4) π is strongly c-monotone.
(5) π is transitively c-monotone.

PROOF. We have already proved that (5) $\Rightarrow$ (4) $\Rightarrow$ (3) $\Rightarrow$ (1) $\Rightarrow$ (2). The implication (2) $\Rightarrow$ (5) is a consequence of Lemma 4.2. $\square$

The following example shows that the ($\mu\otimes\nu$-a.e.) finiteness of the cost function is essential to be able to pass from the "weak properties" (optimality, c-monotonicity) to the "strong properties" (robust optimality, strong c-monotonicity).

Example 5.3 (Optimality does not imply strong c-monotonicity) *Let $X = Y = [0,1]$ and equip both spaces with Lebesgue measure $\lambda = \mu = \nu$. Define c to be ∞ above the diagonal and $1-\sqrt{x-y}$ for $y \le x$. The optimal (in this case the only finite) transport plan is the Lebesgue measure π on the diagonal Δ. We claim that π is not strongly c-monotone. Striving for a contradiction we assume that there exist φ and ψ witnessing the strong c-monotonicity. Let Δ_1 be the full-measure subset of Δ on which $\varphi + \psi = c$, and write $p_X\Delta_1$ for the set $\{x : (x,x) \in \Delta_1\}$. We claim that*

$$\forall x, x' \in p_X\Delta_1 : \textit{ If } x < x', \textit{ then } \varphi(x) - \varphi(x') \ge \sqrt{x'-x}, \tag{26}$$

which will yield a contradiction when combined with the fact that $p_X\Delta_1$ is dense.

Our claim (26) follows directly from

$$\varphi(x') + \psi(x) \le c(x', x) = 1 - \sqrt{x'-x} \quad \textit{and} \quad \varphi(x) + \psi(x) = c(x,x) = 1.$$

Now let $x < x + a$ be elements of $p_X\Delta_1$, let $b := \varphi(x) - \varphi(x')$, and let n be a sufficiently large natural number, say satisfying $n > 2\frac{b^2}{a^2}$. Using the fact that $p_X\Delta_1$ is dense, we can find real numbers $x = x_0 < x_1 < \cdots < x_n = x + a$ in Δ_1 satisfying $x_k - x_{k-1} < 2/n$ for $k = 1, \ldots, n$.

Let $\varepsilon_k := x_k - x_{k-1}$ for $k = 1, \ldots, n$. Then we have $\varepsilon_k < \frac{2}{n} < \frac{a^2}{b^2}$ for all k, hence $\sqrt{\varepsilon_k} > \frac{b}{a}\varepsilon_k$. So we get

$$b = \varphi(x) - \varphi(x') = \sum_{k=1}^{n} \varphi(x_{k-1}) - \varphi(x_k) \ge \sum_{k=1}^{n} \sqrt{\varepsilon_k} > \sum_{k=1}^{n} \frac{b}{a}\varepsilon_k = \frac{b}{a}\sum_{k=1}^{n} \varepsilon_k = b,$$

a contradiction. (By letting $c = 0$ below the diagonal the argument could be simplified, but then we would lose lower semi-continuity of c.)

References

[1] L. Ambrosio, A. Pratelli, Existence and stability results in the L^1 theory of optimal transportation, in: Optimal transportation and applications (Martina

Franca, 2001), vol. 1813 of Lecture Notes in Math., Springer, Berlin, 2003, pp. 123–160.

[2] W. Gangbo, R. J. McCann, The geometry of optimal transportation, Acta Math. 177 (2) (1996) 113–161.

[3] O. Kallenberg, Foundations of modern probability, Probability and its Applications (New York), Springer-Verlag, New York, 1997.

[4] A. S. Kechris, Classical descriptive set theory, vol. 156 of Graduate Texts in Mathematics, Springer-Verlag, New York, 1995.

[5] H. G. Kellerer, Duality theorems for marginal problems, Z. Wahrsch. Verw. Gebiete 67 (4) (1984) 399–432.

[6] M. Knott, C. Smith, On Hoeffding-Fréchet bounds and cyclic monotone relations, J. Multivariate Anal. 40 (2) (1992) 328–334.

[7] A. Pratelli, On the sufficiency of c-cyclical monotonicity for optimality of transport plans, Math. Z.In press.

[8] S. T. Rachev, L. Rüschendorf, Mass transportation problems. Vol. I, Probability and its Applications (New York), Springer-Verlag, New York, 1998.

[9] D. Ramachandran, Perfect measures and related topics, in: Handbook of measure theory, Vol. I, II, North-Holland, Amsterdam, 2002, pp. 765–786.

[10] R. T. Rockafellar, Characterization of the subdifferentials of convex functions, Pacific J. Math. 17 (1966) 497–510.

[11] L. Rüschendorf, On c-optimal random variables, Statist. Probab. Lett. 27 (3) (1996) 267–270.

[12] W. Schachermayer, J. Teichmann, Characterization of optimal transport plans for the Monge-Kantorovich problem., Proceedings of the A.M.S.In press.

[13] C. Villani, Topics in optimal transportation, vol. 58 of Graduate Studies in Mathematics, American Mathematical Society, Providence, RI, 2003.

[14] C. Villani, Optimal transport, old and new, Lecture Notes (2005).